Mat. Res. Soc. Symp. Proc. Vol. 724 © 2002 Materials Research Society N4.6

# Selectivity of Polypeptide Binding to Nanoscale Substrates

Steven R. Lustig and Anand Jagota
Central Research & Development, E.I. du Pont de Nemours & Co., Inc.
Experimental Station, Route 141
Wilmington, DE 19880-0356, U.S.A.

## ABSTRACT

We present new computational methodology for designing polymers, such as polypeptides and polyelectrolytes, which can selectively recognize nanostructured substrates. The methodology applies to polymers which might be used to: control placement and assembly for electronic devices, template structure during materials synthesis, as well as add new biological and chemical functionality to surfaces. Optimization of the polymer configurational sequence permits enhancement of both binding energy on and binding selectivity between one or more atomistic surfaces. A novel Continuous Rotational Isomeric State (CRIS) method permits continuous backbone torsion sampling and is seen to be critical in binding optimization problems where chain flexibility is important. We illustrate selective polypeptide binding between either analytic, uniformly charged surfaces or atomistic GaAs(100), GaAs(110) and GaAs(111) surfaces. Computational results compare very favorably with prior experimental phage display observations [S.R. Whaley *et al.*, *Nature*, **405**, 665 (2000)] for GaAs substrates. Further investigation indicates that chain flexibility is important to exhibit selective binding between surfaces of similar charge density. Such chains begin with sequences which repel the surfaces, continue with sequences that attract the surface and end with sequences that neither attract nor repel strongly.

## INTRODUCTION

We present new computational methodology for designing polymers, such as polypeptides and polyelectrolytes, which can selectively recognize nanostructured substrates. The methodology applies to polymers which might be used to: control placement and assembly for electronic devices, template structure during materials synthesis, as well as add new biological and chemical functionality to surfaces. Optimization of the polymer configurational sequence permits enhancement of both binding energy on and binding selectivity between one or more atomistic surfaces. This optimization is enabled by combining highly-efficient, atomistic modeling of the polymer and surfaces with genetic mutation of the polymer configuration. The atomistic modeling permits the calculation of macromolecular statistics and thermodynamics of substrate binding, while genetic sequence mutation enables the search and enhancement of the desired polymer-surface interactions.

Previous experimental works have demonstrated polypeptides with selectivity for binding to surfaces of metals and metal oxides [1-8] as well as a range of semiconductor surfaces [9]. Polypeptides which can recognize desired surfaces are typically selected from a library of several million candidates using either bare proteins or phages, often in the presence of surfactants or salts. These methods are often both practical and useful. There still exist several issues. First, it is not always clear whether the selected polypeptides will retain their binding and selectivity once

removed from the parent protein or phage body. Second, practical experimental libraries of even $10^8$ candidates might not well represent the complete range of functionalities present in the $>10^{15}$ possibilities from natural residues. Third, experimental screening does not typically teach why particular consensus sequences emerge. Hence we might not always be able to predict new and better binding sequences. Finally, we ask if it is possible to design better polymer sequences and compositions than those available from natural sources and residues.

If theoretical and computational methods are to be as practical and useful, they will surely need to contain the salient physics and chemistries of polymers and surfaces while remaining both accurate and quickly solvable. Toward this end, we illustrate methodology for selective polypeptide binding between either analytic, uniformly-charged surfaces or atomistic model surfaces. Here we compare our preliminary findings to recent pioneering, experimental results [9]. Further, we ask how to find optimal sequences which selective a target surface over closely similar surfaces.

## COMPUTATIONAL METHODS

Polypeptides are described as rotational isomeric state chains in which bond lengths and bond angles are frozen at equilibrium values while torsional rotations remain degrees of freedom. In applying discrete Rotational Isomeric State, RIS, theory [see e.g. 10-12], we select discrete energy states at the minima in a potential energy surface from mapping pairwise-conditional, rotational angles around neighboring N-$C_\alpha$ and $C_\alpha$-C' bonds for each natural amino acid residue. In applying our new Continuous Rotational Isomeric State, CRIS, method, the torsional angles may be selected within a continuous range from rectangular tiles around minima in the potential energy surface. The tiles are defined from energy minima bounded by a preset, maximum well height, typically 1 kcal/mol. If a local potential energy surface exhibits a relative maximum before reaching the preset height, then the tile boundary is defined at the relative maximum. Tile boundaries are combined by overall union if the tile definitions create overlapping regions from multiple energy minima. A chain backbone conformation is completely defined from the fixed bond lengths, fixed bond angles and selected torsions. Non-backbone atom positions are described as "pendtors", i.e. pendant vectors. Amide hydrogens and oxygens are placed on the backbone from constant vectorial components using a basis set generated from the [C'-N], [N-$C_\alpha$] bond vectors and their cross-product. Likewise each residue's pendant atoms are placed from constant vectorial components using a basis set generated from the [N-$C_\alpha$], [$C_\alpha$ -C'] bond vectors and their cross-product. The various rotational states within each residue's pendant group can be represented by sets of their atomic pendtors. A rotational potential energy surface and the aforementioned fixed geometrical parameters for each amino acid residue with amide terminations was created using the PCFF forcefield [13,14]. When implementing RIS, a table of discrete rotational states (each state comprising the conditional pair of torsion angles and the associated energy value) for each residue is stored in memory during a simulation to look-up the rotational energy. When implementing CRIS the entire pairwise-conditional torsional energy surface for each residue is simply stored in memory during a simulation to look-up and interpolate the rotational energy.

The polypeptide potential energy includes additional interatomic contributions. Self-avoidance is ensured by assigning hard-sphere radius, typically 0.5Å, for each atom. The hydrophobic effect is approximated by contributing a fixed energy decrement, typically –0.25

kcal/mol, when two hydrophobic groups reach a minimum separation, typically 5Å. Electrostatic potential energy between atoms arises from partial charges for each atom assigned by the COMPASS forcefield [15]. Since the polypeptide is ensconced in an effective solvent medium, atoms experience diminished electrostatic potentials, $V$, through the Debye-Hückel potential,

$$V = \left(\frac{L e^2}{4\pi\varepsilon_o}\right)\left(\frac{q_i q_j}{\varepsilon r_{ij}}\right)\exp(-\kappa r_{ij}) \tag{1}$$

where the first enclosed term is the thermal Bjerrum length, $q_i$ is the partial charge on the ith atom, $\varepsilon$ is the solvent dielectric constant, $\kappa$ is the inverse electrostatic screening or Debye length which is dictated in actual experiments by the concentration of dissolved ions and $r_{ij}$ is the distance between the ith and Jth atoms. Equation (1) applies also between polymer and surface atoms. For infinite analytic surfaces with uniform charge, we compute the potential between atoms and the analytic surface using the integrated form of Equation (1) below:

$$V = \left(\frac{L e^2}{2\varepsilon_o}\right)\left(\frac{\sigma q_i}{\varepsilon\kappa}\right)\exp(-\kappa z_i) \tag{2}$$

where $\sigma$ is the surface charge density and $z_i$ is the height above the surface of the ith atom.

Our simulations utilize an unusual methodology to sample the polymer degrees of freedom (sometimes referred as "simple sampling Monte Carlo" [16] or direct phase-space integration) for computing molecular statistics and thermodynamics. The degrees of freedom consist of: the internal torsion states along the polymer backbone, the position of an end bond vector relative to the surface origin and a rigid-body rotation of the chain around the end bond vector. Torsional states are selected randomly such that each discrete state (for RIS) or position in a torsional tile state (for CRIS) is selectable with equal and uniform probability. Phage display and tumble chain sampling methods are used to integrate over the remaining spatial degrees of freedom. Phage display sampling always assigns the C'-terminus bond vector normal to the surface (with the penultimate bond vector toward the surface) and choses a random, rigid-body rotation about the end bond vector. Note that the polypeptide C'-terminus is fixed so that the N-terminus can be displayed to the surface to emulate a phage peptide display. Tumble sampling comprises selecting a random spatial orientation of an end bond vector and selecting a random rigid-body rotation about the end bond vector. For both sampling methods the absolute distance between the lowest atom in the polymer and the highest component of the surface is varied to sample a profile of important statistical quantities as a function of height above the surface. Since all Monte Carlo trials are always "accepted", statistical quantities are computed with each configuration being weighted by its appropriate thermal Boltzmann factor, $\exp[-(E-E_{form})/kT]$. Here E is the total potential energy of a Monte Carlo trial and $E_{form}$ is the energy of formation. Statistical quantities of interest include the well known polymer-surface binding free energy, A, internal energy, U, entropy, S, binding constant, K, as well as geometrical shape changes, e.g. strain, $\varepsilon$, and squash, $\xi$.

$$\epsilon = \frac{1}{2}\left(\frac{\langle R^2\rangle_{surface}}{\langle R^2\rangle_{solvent}} - 1\right), \qquad \xi = \left(\frac{\langle R^2_{\Leftrightarrow}\rangle_{surface} - \langle R^2_{\Updownarrow}\rangle_{surface}}{\langle R^2\rangle_{solvent}}\right) \tag{3a,b}$$

Here we define a strain in terms of the mean-squared end-to-end chain distance on the surface, $\langle R^2 \rangle_{surface}$, relative to free solvent, $\langle R^2 \rangle_{solvent}$. The chain squash is defined in terms of the mean-squared, chain end-to-end vector components parallel to the surface (designated by horizontal arrows) and normal to the surface (designated by vertical arrows).

## RESULTS

Figure 1 illustrates the typical convergence of a polypeptide-surface binding free energy as a function of the number of Monte Carlo samples. This example polypeptide has over $10^7$

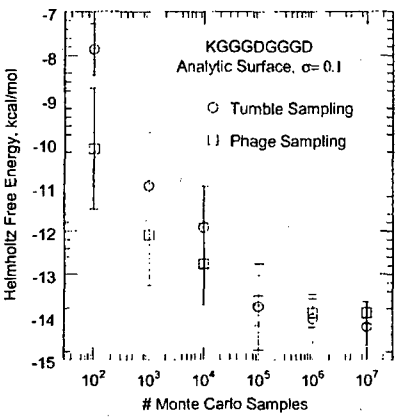


**Figure 1.** Polypeptide-surface binding free energy as a function of Monte Carlo samples using CRIS chain model. The test polypeptide is presented a flat surface with uniform charge density 0.1 e/Å$^2$. Error bars indicate the standard deviation from 10 independent, replicated simulations.

discrete torsional degrees of freedom yet the free energy has less than 1.5 kcal/mol uncertainty after $10^5$ phage samples and $10^6$ tumble samples. Note that this particular chain is a very strong binder. Metropolis Monte Carlo methods typically do not converge as quickly [17] and require extensive equilibration to surmount chain conformation trapping in deep potential energy wells which is intrinsic to these systems [18].

Our computational results compare favorably with previously-published, experimental phage display observations. We subjected model RIS chains to phage sampling over an atomistic GaAs(100) surface model. The residue sequences match the pIII coat proteins of M13 coliphages reported in Figure 1 of reference [9] which were found to bind to GaAs(100) substrates. While these preliminary computations predict 8 of the 11 polypeptides have very favorable binding free energies ($\Delta A < 0$), all chains are predicted to bind significantly to the surface. The statistical binding constant, K, takes a value of unity for permanent surface binding and zero for no surface binding. Our computed values of K range from 0.4 to 0.9, which is consistent with elutable surface binding. Typically, we find that those chains with the lowest binding free energies also show the greatest tendency to spread across the surface (high values of $\varepsilon$ and $\xi$). Most significantly, our computations do reproduce experimental observations that clone G1-3 exhibits preferential binding to GaAs(100), over both gallium and arsenic terminated GaAs(111) faces.

**TABLE I.** Binding results for RIS chain models of phage display polypeptides (see Figure 1 of reference [9]) tested by phage sampling over atomic GaAs(100) model.

| Phage [9] | ΔA | ΔU | $\Delta S x 10^3$ | K | ε | ξ |
|---|---|---|---|---|---|---|
| G1-3 | -1.6 | -1.9 | -0.9 | 0.9 | 2.19 | 0.45 |
| G1-4 | -0.6 | 0.5 | 3.8 | 0.7 | 1.64 | -1.80 |
| G7-4 | 0.3 | -0.6 | -3.1 | 0.4 | -0.19 | -0.10 |
| G11-3 | 0.2 | 1.0 | 2.6 | 0.4 | -0.16 | -0.17 |
| G12-3 | -0.9 | 1.3 | 7.3 | 0.8 | 0.26 | 0.94 |
| G12-4 | -0.5 | -1.2 | -5.6 | 0.7 | -0.18 | 0.10 |
| G12-5 | 0.3 | -0.6 | -2.8 | 0.4 | -0.27 | -0.06 |
| G13-5 | -1.0 | 1.1 | 7.0 | 0.8 | 0.58 | 1.66 |
| G14-3 | -1.0 | 0.8 | 6.0 | 0.9 | 1.62 | 0.93 |
| G14-4 | -0.4 | -0.2 | 0.9 | 0.7 | 0.28 | 1.41 |
| G15-5 | -0.8 | 0.2 | 3.5 | 0.8 | 0.20 | -0.14 |

In a separate set of computational experiments, we explore how to construct chains to bind selectively to only modestly, Lewis acidic surfaces. Sequences were limited to the residue-pair combinations of: KP (stiff base), KG (flexible base), DP (stiff acid), DG (flexible acid), GG (flexible neutral) and PP (stiff neutral). Furthermore we compare fully atomistic chain models to united-atom chain models (where each backbone entity possesses the net charge from all of its pendant atoms while preserving RIS properties for each residue). Not surprisingly, we find chains with high base residue content bind strongly to surfaces with increasing surface charge density, but this is only trivial selectivity. More interestingly, we found only chains with flexible residue pairs exhibit non-monotonic binding to surfaces with increasing surface charge density. Table II illustrates sequences with optimized binding constants for intermediate-valued surface charge densities. Such chains begin with base sequences which repel the surfaces, continue with acid sequences that attract the surface and end with neutral sequences that neither attract nor repel strongly. This suggest that several types of star polymers could exhibit interesting selective binding properties.

**TABLE II.** Surface binding constant, K, for RIS united atom chain models on analytic surfaces.

| | Surface Charge Density ($e/Å^2$) | | | | |
|---|---|---|---|---|---|
| **Sequence** | **0.20** | **0.16** | **0.12** | **0.08** | **0.04** |
| $(KP)_2 (DP)_2 (G)_{12}$ | 0.90 | 0.98 | 0.82 | 0.48 | 0.51 |
| $(KG)_2 (DG)_2 (G)_{12}$ | 0.92 | 1.00 | 0.69 | 0.51 | 0.52 |
| $(KG)_2 (DP)_2 (G)_{12}$ | 0.96 | 0.96 | 0.46 | 0.55 | 0.51 |
| $(KG)_2 (DG)_2 (P)_{12}$ | 0.98 | 0.74 | 0.86 | 0.63 | 0.56 |


## ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank Angela Belcher (Massachusetts Institute of Technology) and Siqun Wang (E.I. du Pont de Nemours & Co., Inc.) for very helpful discussion.

Correspondence should be addressed to S.R.L. (Email at the time or writing this manuscript: steve.r.lustig@usa.dupont.com)

**REFERENCES**

1. W.D. Huse, L. Sastry, S.A. Iverson, A.S. Kang, M. Alting-Mees, D.R. Burton, S.J. Benkovic and R.A. Lerner, *Science*, **246**, 1275-1281 (1989).
2. J.D. Marks, H.R. Hoogboom, T.P. Bonnert, J. McCafferty, A.D. Griffiths and G. Winter, *J. Mol. Biol.* **222**, 581-597 (1991).
3. J.K. Scott and G.P. Smith, *Science*, **249**, 386-390 (1990).
4. S.E. Cwirla, E.A. Peters, R.W. Barrett and W.J. Dower, *Proc. Natl. Acad. Sci. USA*, **87**, 6378-6382 (1990).
5. J.J. Devlin, L.C. Panganiban, P.E. Devlin, *Science*, **249**, 404-406 (1990).
6. K.S. Lam, S.E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski and R.J. Knapp, *Nature (London)*, **354**, 82-84 (1991).
7. S. Brown, *Proc. Natl. Acad. Sci. USA*, **89**, 8651-8655 (1992).
8. S. Brown, *Nature Biotechnol. 15, 269-272 (1997).*
9. S.R. Whaley, D.S. English, E.L. Hu, P.F. Barbara and A.M. Belcher, *Nature*, **405**, 665-668 (2000).
10. P.J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).
11. W.L. Mattice and U.W. Suter, *Conformational Theory of Large Molecules; The Rotational Isomeric State Model in Macromolecular Systems* (Wiley, New York, 1994).
12. M. Rehahn, W.L. Mattice and U.W. Suter, "Rotational Isomeric State Models in Macromolecular Systems," *Advances in Polymer Science*, **131/132**, (Springer, New York, 1997).
13. J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comput. Chem.* **15**, 162-182 (1994)
14. H. Sun, *J. Comput. Chem. 15*, 752-757 (1994)
15. At the time of writing this manuscript, the COMPASS force field is a licensed product of Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121 (http://www.accelrys.com). See for example: H. Sun and D. Rigby, *Spectrochim. Acta*, **53A**, 1301 (1997); H. Sun, *J. Phys. Chem. B*, **102**, 7338-7364 (1998); H. Sun, P. Ren and J. R. Fried, *Comput. Theor. Polym. Sci*, **8**, 229. (1998).
16. K. Binder and D.W. Heermann, "Monte Carlo Simulation in Statistical Physics", 2nd corrected ed., *Solid-State Sciences*, **80**, (Springer-Verlag, New York, 1992).
17. See for example: C.Y. Kong and M. Muthukumar, *J. Chem. Phys.* **109**, 1552-1527 (1998); Y.-H. Lee and B.J. Berne, *J. Phys. Chem.*, **104**, 86-95 (2000); S. Santos, U.W. Suter, M Muller and J. Nievergelt, *J. Chem. Phys.* **114**, 9772-9779 (2001); P. Chodanowski and S. Stoll, *Macromolecules*, **34**, 2320-2328 (2001).
18. M Muthukumar, *Proc. Nat. Acad. Sci. USA.* **96**, 11690-11692 (1999).